Behavioral Observation Exposure Study Plan

The U.S. Consumer Product Safety Commission (CPSC) staff is investigating the potential exposure and health risks to children under three years of age from teethers, rattles, and toys made from polyvinyl chloride (PVC) that contains various dialkyl phthalate (DAP) plasticizers, especially diisononyl phthalate (DINP). Manufacturers use plasticizers to soften the PVC.

The potential for DINP to cause toxic effects in humans depends on the amount of DINP that is released from a product when it is mouthed or chewed and the amount of time a child spends putting that product in his or her mouth. In December 1998, CPSC staff released the results of a study of DINP in children's products.[1] The staff concluded that based on the best available information, few, if any, children are at risk of liver or other organ toxicity from mouthing PVC toys that contain DINP. This conclusion is based on estimates of the amount of DINP ingested, which indicated that DINP exposure did not reach a potentially harmful level. However, the staff believes that there are a number of uncertainties in this assessment, particularly regarding the types of toys that children are mouthing and how long these toys are typically in their mouths. Staff will undertake additional work to gather better data on which to base the health risk assessment. One aspect of this additional work is to conduct an extensive exposure study to obtain data that will better define the amount of time that children mouth products.

A previous study conducted by a Dutch Consensus Group observed mouthing behaviors of children of ages between 3 and 27 months.[2] In this study, Dutch parents were the observers, and they recorded their child's mouthing behavior during ten 15-minute periods over 2 days in the home. The results indicated that the youngest children perform mouthing behaviors more often and for a longer total time than older children. However, based on the limited data on older children, it is not clear when this mouthing behavior subsides. In addition, the data from this study were only collected from the child's home environment, thereby raising questions about children's mouthing behaviors in different environments and under different circumstances.

Objective

The objective of this study is to quantify the time per day that young children spend mouthing objects, including toys and other children's products, and especially those containing phthalates.

---

[1] U.S. Consumer Product Safety Commission (CPSC, 1998). "The Risk of Chronic Toxicity Associated with Exposure to Diisononyl Phthalate (DINP) in Children's Products." Washington, DC.

[2] Groot, M.E., Lekkerkerk, M.C., and Steenbekkers, L.P.A., (1998). "Mouthing Behavior of Young Children." Agricultural University Wageningen, Wageningen, The Netherlands.

Method

Subjects

A total of two hundred children from 3 through 36 months of age will be observed in two different geographical areas (100 children in each area). The only formal quotas for subjects will be age and area. For each area, there will be 40 children from 3 through12 months, 30 children from 13 through 24 months, and 30 children from 25 through 36 months. To ensure that the subjects will be reasonably representative of the population with regard to race, income, type of child care, and gender, the two sites will have a large enough population so that there will be large subpopulations in each racial and income group and in different types of child care.

During the sampling process, the children recruited for the study will be monitored for gender, age, income, race, and child-care groups. Quota sampling may be imposed if the samples show substantial underrepresentation of any of these groups. Using Random Digit Dialing techniques, children meeting the criteria will be identified in two Metropolitan Statistical Areas (MSA); for example, Raleigh-Durham, Detroit, or Boston. The observational study contractors will propose areas to be studied. There will be a mailing of additional information to the observational study families in order to encourage them to participate in the study. The principal investigators from each contractor will also be involved in the recruitment.

Observations

The exposure study will consist of a series of half-day observations in typical settings (e.g., home, babysitter, childcare, etc.). The observer will travel with the child between places, when necessary. The observer will not introduce anything new to the setting(s), except himself or herself. The observer will not communicate with the child aside from that necessary for initial habituation.

Before the observations are conducted, the observer will meet with the parent(s) and the child. This will familiarize the parent(s) and child with the observer, and decrease the novelty effect when the observations are initiated. At this initial meeting, the observer will ask the parent(s) to provide information about the child's normal routine of waking, sleeping, and eating times, daycare time and location, and favorite toys. The observations will take place for four hours in the morning and four hours in the afternoon/evening. If the child is cared for outside of the home, the observations will be arranged so both home and day care locations are observed. The observer will keep a diary of the child's activities during the time the observer is present. This will create a richer record of the child's typical activities and environments, which may be useful for post hoc analysis.

For 10 or 15 minutes out of each half-hour, the child's mouthing activities will be recorded. The start time for the 10- or 15-minute period will be randomly generated for each half-hour. For purposes of this study, mouthing is defined as any behavior other than eating or drinking in which an item comes into contact with the observed child's lips, tongue, or inside of the mouth for any length of time. The item being mouthed will be

identified as precisely as possible. The observer will use a stopwatch and data collection sheet to record the starting time, stopping time, type of mouthing behavior, and type of item mouthed. If a daycare facility is equipped with video recording materials, some of the children will be videotaped for visual documentation of children's behaviors. The staff does not expect to routinely use videotape recording, however, because it may introduce a distraction into the environment.

Phone Survey

There will be 400 telephone interviews (200 in each area) of families with children 37 through 72 months of age. The objective of these is to estimate the proportion of children over 36 months old who are mouthing toys and other objects. Based on the results from the Dutch study, CPSC staff expect mouthing behaviors to decrease with age. Therefore, it would not be productive to observe children who may no longer be mouthing toys and other objects. However, if the phone survey reveals that mouthing behavior is still common beyond 36 months of age, timed observations of this behavior can be done during a later research phase. At the beginning of the interviews, parents will be asked to participate in the study. If they agree, they will be mailed a questionnaire and instructed on how to observe their children over a few days. A few days later the Contractor will phone them to collect their answers. The phone survey participants will be recruited during the screening for participants for the observational study. No telephone interview will exceed 15 minutes in length.

Contractors

One contractor will be hired to conduct the phone recruitment and the older children's survey. Another contractor will be responsible for creating the training materials. One or two additional contractors will conduct the observations. The observation contractor(s) will be responsible for training (using CPSC-provided training materials) and providing an adequate number of observers to complete the observations within the time specified by CPSC staff. All observers will be trained using the same materials, part of which will be videotaped observations of children. All observers will be required to watch and record the behaviors of the children on the tapes. The data will then be compared for inter-observer reliability.

The observation contractor(s) will be selected based on location, experience with child observational studies, ability to prove competently and adequately conduct observations, and ability to accurately collect the data. CPSC staff will monitor data collection and be responsible for the analysis.

Data Analysis

1. Identify the statistical distribution (i.e. parametric model) of the mouthing duration by age group (3-12 months, 13-24 months and 25-36 months), or determine that no distribution provides an adequate fit to the data. Conduct a statistical test on the goodness of fit of the statistical distributions to the data.

2. Determine whether variability in the distribution is associated with income, gender or other variables. This might lead to a stratified analysis of one or more of these variables.
3. From the results of tasks 1 and 2, estimate the relevant parameters of the mouthing time distribution, if a parametric model is appropriate.
4. Analyze the telephone survey data to determine whether children over 36 months perform mouthing behaviors in sufficient amounts to require timed observations.
5. Develop a model for estimating DINP intake using the appropriate parametric model, parametric bootstrap model, or bootstrap model, depending on the outcome of the data analysis.

Further details of these tasks are found in Appendix A.

Peer Review

The Commission will solicit the assistance of several peers to review and comment on the study plan. The reviewers will be used during the developmental stage of the plan and before the documentation is submitted to OMB for clearance.

CPSC Pilot Study

Commission staff will conduct a pilot study with staff volunteers to refine the procedures. No more than nine children will be observed during this pilot study. Issues that will be monitored include observer fatigue, logistics of transportation to and from the child's location, questionnaire design, and the child's habituation to the observer. These sessions may also be videotaped for use in the training materials. This phase is only preliminary, and no data derived from these initial sessions will be used in the primary study. Staff will incorporate the refined procedures into contract specifications for the observation contractors.

Summary

This study is initiated under the Phthalates Project because we are concerned with children's exposure to phthalates in their toys and other products. However, the results of this study may allow Commission staff to calculate exposures to other potential hazards. This study will give the CPSC staff quantified data of children's overall mouthing behaviors and allow the staff to more accurately determine which items in a child's environment are hazardous.

Appendix A
Statistical Background for the Behavioral Observation Study

This appendix presents details on three issues in the study plan. These are (1) the sample size for the observational study of children's mouthing behavior, (2) the telephone survey and (3) the data analysis plan. The previous research, which forms the basis for these proposals, is discussed first.

Previous Research

The basis for the sample size calculations and the data analysis plan are the previous research on mouthing and DINP intake by the Dutch Consensus Group (Konemann et al, 1998) and the U. S. Consumer Product Safety Commission's statistical analysis of DINP exposure (Babich, 1998; Greene, 1998). Both the Dutch Consensus Group (DCG) and the Consumer Product Safety Commission (CPSC) used observational data from a sample of Dutch children aged 3-36 months (Groot, Lekerkerk and Steenbekkers, 1998). These data on mouthing durations will be referred to in this paper as the "Dutch data."

DINP Intake Model

The mouthing distributions were a component of the model for DINP intake. The mathematical model for estimation of chronic toxicity was presented in Babich (1998, page 10). Estimated oral exposure was given by equation (1) as

$$E = MDk / W \qquad (1)$$

where

E = Exposure
M = DINP Migration Rate
D = Mouthing Duration
k = constants
W = body weight

Organ toxicity and cancer risk are a function of exposure, E, which is in turn a function of the two key distributions, namely those of M, migration rates and D, mouthing duration.

Both the CPSC and DCG used equation (1) to estimate the upper percentiles of the exposure distribution. The DCG used a Monte Carlo or (nonparametric) bootstrap simulation procedure for the upper percentiles. This procedure involved creating the distribution of exposure from repeatedly resampling from the duration and migration rate data (also known as empirical

distributions). A single sample from the data would contain a migration rate from a given object, *M*, and a mouthing duration, *D*. When multiplied together and suitably scaled with the proper constants, this would provide a single observation for exposure, *E*. The distribution of exposure, *E*, can then be built up with repeated resampling. Since the resampling process is meant to represent the DINP intake in the population, the distribution of *E* should trace out the actual exposure distribution in the population. The upper percentiles are then available from the tail of the distribution. In other words, the 95th percentile of the simulated exposure distribution, *E*, would be an estimate for the 95th percentile exposure to DINP in the population.

The CPSC approach was somewhat different. The CPSC was able to find a closed form expression (i.e. a mathematical formula) for the distribution of exposure. Upper percentiles were computed in closed form. Confidence intervals for the percentiles were obtained initially from tables (Hahn and Meeker, 1991) and using the parametric bootstrap. The parametric bootstrap differs from the procedure used by the DCG because the sample is from a mathematical distribution instead of the empirical distribution. This is done with statistical software that has built-in random number generators that reflect various theoretical distributions.

Which approach is better, using a simulation to sample from the empirical distribution or using a mathematical model? Many statisticians believe that if a model provides a good fit to the data, and the model is plausible, then the model is better. Working with a model is computationally simpler, and values can be compared with statistical tables. This eliminates the possibility of some programming errors. On the other hand, if the model does not fit the data well, then the empirical distribution approach is better.

Data and Data Analysis

Data for estimating the distribution of migration rates, *M*, were obtained from in-vitro impaction experiments from 31 products carried out at the CPSC Directorate for Laboratory Sciences (Chen, 1998). These data were scaled by the ratio of in-vivo to in-vitro extraction migration rates to provide estimates for the migration rate per unit time that would be experienced by a human subject. Both CPSC and the DCG used the same mouthing duration, *D*, data. Body weight, *W*, was taken from anthropometric data in Snyder, et al (1977) and treated as a constant.

In both CPSC and DCG analyses, it was shown that mouthing durations for toys was highly skewed with a long right hand tail. The CPSC analysis fit the data with a lognormal distribution. Separate models were fit for children 3-12 months old and 13-26 months old.[1] The log transformed data passed the Shapiro-Wilk test for the normal distribution. The underlying normal distributions had mean and variance (2.49 and 1.87 respectively) for children age 3-12 months and (0.75 and 1.12 respectively) for children 13-26 months.

[1] The age group includes children whose age on the next monthly birthday will be at or within within the endpoints of the age interval. Children 3-12 months old will be no younger than 3 months on their next monthly birthday and no older than 12 months.

Analysis of the distribution of CPSC migration rates, *M*, also showed that it could be fit with a lognormal distribution. All data were transformed to logs, parameters were estimated and the distribution of the log of exposure, log(*E*), was obtained by summing the parameters for the distribution of the log of duration, log(*D*) and the log of migration rates, log(*M*). All further calculations were then carried out in logarithmic form, and ultimately transformed back. For example, the upper 95th percentile of the distribution of exposure was given as

$$E_\alpha = \exp(m + d + z_\alpha \sqrt{s_m^2 + s_d^2}) \qquad (2)$$

where

$E_\alpha = 100\alpha$ percentile of the distribution of E.
m = estimated mean of the log of migration rates
d = estimated mean of the log of exposure
s^2 = variance of migration rate (subscript m), or duration, (d)

These data and the approach have been used in planning the proposed observational study.

Sample Sizes for the Observational Study

To compute sample sizes the two issues addressed were

- how many children in total should there be in the study and
- how should the children be allocated between age groups 3-12 months, 13-24 months and 25 to 36 months.

The approach taken was that the children should be allocated between the age groups so as to equalize the uncertainty in the estimates of quantities such as upper percentiles of the exposure distribution. This would result in confidence intervals of approximately equal size. The total sample size should be chosen so that the confidence intervals are no wider than of practical importance.

Allocating the number of Children Between Age Groups

Regardless of the mathematical distribution fit, the width of the confidence intervals for the upper percentiles will be a function of several quantities including the standard error of the mean for the log of migration rates, $s_m / \sqrt{n_m}$ and the standard error of the mean for the log of duration, $s_d / \sqrt{n_d}$. (The sample sizes in these quantities are represented by n while s symbolizes the standard deviations.) The approach taken for sample size allocation between the age groups was to

select sample sizes such that the standard error of the duration distribution was the same for all three age groups of 3-12 months, 13-24 months and 25-36 months. To do this, estimated values of the standard deviation from the Dutch data were used. It can be shown that the width of the confidence intervals are equalized when the sample sizes are chosen in proportion to the estimated variances.

Using the variances from the Dutch data (see the previous section), standard errors would be equalized with (1.87/1.12 =) 1.67 children age 3-12 months to 1 child age 13-24 months. Similar calculations could not be made for children 25 to 36 months because in this age group there were only two children in the Dutch data, (one age 26 months and the other 35 months), which is not enough data to reliably estimate the standard errors. We assumed that 25-36 month old children would have about the same variance in mouthing durations as the 13-24 month old children. As a result, the allocation ratio was 1.67:1:1, with the largest sample size allocated to the oldest children.

Total Number of Children

With this allocation scheme, confidence intervals for the 95th percentile daily exposure and for the percentage of children over 150 micrograms per day were calculated. The analysis used the same parameters and distributions as those observed in the Dutch study, and the CPSC 1998 analysis but varied the number of children in the study. The children were allocated according to the 1.67:1 rule. As the total sample size increased, confidence interval widths in both age groups (3-12 and 13-24 months) decreased, but after about sample sizes of 80 children were simulated, the decreases were not of practical importance. For example, the confidence intervals for the 95th percentile exposure with 80 children in the 3-12 month age group was 46.2 to 188.9 around a point estimate of 94.3 micrograms per average body weight per day. With a sample size of 100 this was 47.5 to 186.1, while a sample size of 120 produced 48.3 to 181.1.

Thus, larger sample sizes did not provide a meaningful decrease in the width of these intervals. To some extent this is a result of the width of the exposure confidence intervals also being driven by the standard error in the migration rate distribution (the other standard error term in the previous section).

With 80 children in the 3-12 month age group, the 1.67:1 sample size allocation ratio would imply (80/1.6=) 48 children aged 13-24 months and 48 25-36 months. This was raised to 60 children for the two older age groups to provide additional protection against problems that might be expected from small samples and skewed distributions. This will result in a total of 200 subjects will be observed in two locations (80 children 3-12 months, 60 children 13-24 months and 60 children 25-36 months).

The Telephone Survey

In addition to the observational study, a telephone survey is planned for families with children 37 to 72 months. One purpose of the interviews is to determine the proportion of these children by age that are still mouthing toys.

Families recruited for the phone interview will be selected from the RDD survey as a by-product of recruitment of the children for the observational study. Some of the respondents to the phone survey will have children only between 3-6 years of age. At the present time it is not known whether additional phone calls will be need beyond those to recruit children for the observational study.

Whether 400 interviews will be sufficient or not to estimate the distribution of mouthing by age, depends on how much variability there is in the data. Generally mouthing decreases with age, as shown in the Dutch data but little is known about how much variability there is and what factors are associated with that variability. In a worst case situation, if the distribution does not look like it can be modelled as a function of age, separate estimates will be made by age group for the proportion of children in that age group mouthing toys. With three age categories (3-4, 4-5 and 5-6), and about 133 children per age group, we would expect a standard error of at most 4% for each age group. That is, in the worst case, the 95th percent confidence interval for the proportion would be the point estimate plus or minus 8%.

Data Analysis Plan

As noted above, the primary objective of the study is to estimate the time typical American children spend mouthing toys and other products, especially those containing phthalates. These estimated mouthing times will be used in a mathematical model for estimation of the phthalate intake of children between age 3 and 36 months. This will either be the closed form model used in the CPSC research, or a bootstrap model similar to the DCG or something in between.

Also discussed above, the proportion of children mouthing toys by age groups will be estimated from the telephone survey for children between 37 and 72 months.

From those quantities, it should be possible to make estimates useful to determining chronic organ toxicity and, if warranted, cancer risk. Chronic organ toxicity estimates use the period of highest exposure. This will probably be for children under 12 months, if the patterns are the same as found in the Dutch data. Cancer risk estimates require estimates of cumulative dosage. All the data from the observational study and the telephone survey will be used for the cancer risk estimates.

The tasks involved in analysis of the data include identifying the mathematical distributions and then determining whether the data can be stratified by other variables collected in the study, such as gender or income. Stratification may lower the variance of the mouthing duration data if the durations are very different between data. Once the strata and distributional form are identified, the appropriate statistical technique can be used to estimate parameters for the distribution. It

might also turn out that no distribution can be found which fits the data. This would lead to the type of bootstrap analysis like the DCG used. In the analysis plan, there also will be an attempt to correct estimates for non response, providing some pattern can be found among the non-respondents.

After the estimates are available, the data can be used in a model for analysis and prediction of phthalate intake.

Task 1. Estimate the distribution of mouthing duration

The first step is to determine the mathematical form of the distribution of the duration data, should such a distribution exist. This would be done graphically using quantile-quantile (or QQ) plots where the quantiles of the empirical (observed) distributions are plotted against the quantiles of possible theoretical distributions. These plots are designed so that a good fit is indicated by a straight line. Following the appearance of good fits to the data, statistical tests will be used to confirm the fit. These will include parametric tests that are specific to certain distributions (for example the Shapiro-Wilk test for the normal distributions) and-non parametric tests such as Kolmogorov-Smirnov, that can be used to test any distribution. Usually parametric tests are more powerful, in that they are better at indicating a lack of fit between empirical and theoretical distributions.

The analysis would be stratified by age group, most likely by year of age.

This part of the analysis will have a major impact on the remainder of the project. If a mathematical distribution can be fit that can combine with migration rates to produce an estimate for DINP exposure (E) in closed form, such as the lognormal distribution that was used in the previous study, then such a distribution will be used. If the duration distribution does not seem to fit a lognormal distribution, then it is most likely that the exposure model (1), will be approached in a manner similar to the DCG. If a distribution can be fit at all but not easily combined with the migration rate distribution, then parametric bootstrap analysis will be used for the exposure distribution.

Although stratified by age, the data reflects a mixture of mouthing durations from children of different genders and different areas of the country. Whether further stratification beyond age is useful would be considered in task 2.

Task 2. Assess the variability in the distribution of the mouthing distributions

There are several reasons for stratifying the data. First, should mouthing durations vary by strata, then controlling for stratification may reduce the variability of the duration distribution. Second, if there is sizable nonresponse by stratum, it may be important to weight the respondents for nonresponse. See task 3 below.

Possible strata could include gender, area where the sample study was conducted, family income, time of day when the observations were made, or day care arrangements.

This analysis would be somewhat different from task 1, because the primary comparison would be between empirical distributions in this task, rather than between an empirical distribution and a theoretical distribution. The same tools would be used for comparisons of empirical distributions, namely quantile-quantile plots, and non parametric tests to determine if both distributions can be said to come from the same population. Also, if both strata appear to fit certain theoretical distributions, parametric tests may be used to determine if there are differences between strata.

Task 3. Correction for nonresponse.

It is expected that some people with children between 3 and 36 months will refuse to participate in the observational study. We propose to collect at least some of the following data from all phone call recipients who have a child under 6 years of age: family income, child care arrangements, child age and gender. Should the stratification analysis prove to be important in Task 2, then the mouthing durations from the observed children will be weighted for nonresponse. Weights are the inverse of the selection probability which in this case would be the number of people who have children 3-36 months old and participate divided by the number of participants plus refusals.

Task 4. Analysis of the phone survey reports on older children's mouthing activity.

In this task we expect data from the phone survey that would provide the age, and gender of up to 400 children between 3 and 6 years old and whether they are currently mouthing toys.

From these data, we intend to model the probability that a child is still mouthing toys as a function of age, and any other relevant variable that can be obtained from the phone survey. A monotonic response with age, probably that mouthing activity is decreasing with age would be expected. If no such relationship can be found, then the analysis will simply estimate the percent of children mouthing toys by age group and the associated standard errors. This was discussed previously in the section about the sample size calculations for the phone study.

This task is primarily relevant to long term exposure to DINP, rather than the dosage during the highest exposure period because it is unlikely that children over three years of age have substantial mouthing activity. Should the proportion of children mouthing toys in the older age group be large, however, it may be necessary to extend the observational study to some or all of the children in this age group. This would be considered in a subsequent research project

Task 5. Estimate the parameters of the mouthing distribution.

Following completion of the distributional analysis, and if a parametric distribution can be identified, estimates will be made of the parameters of the distribution using the best available statistical methods. If a parametric distribution cannot be identified from tasks 1 and 2, this task will be omitted.

Task 6. Analysis and prediction of DINP intake.

DINP intake will be estimated using the appropriate parametric, parametric bootstrap or non parametric boostrap method. If a parametric model can be used, the analysis will be similar to Greene (1998) . Otherwise, the research will follow the non parametric bootstrap method used by the DCG (1998).

At the conclusion of this analysis, there should be new estimates for DINP intake, and if desired, the cancer risk. If there appear to be important strata as discussed in task 2, it may be possible to produce stratified estimates. The larger sample size than the previous study, should permit smaller confidence intervals and more precision in the predictions.

References

Babich, M. (1998), "The Risk of Chronic Toxicity Associated with Exposure to Diisononyl Phthalate (DINP) in Children's Products", U.S. Consumer Product Safety Commission. Bethesda, MD.

Chen, S-B, (1998), "Migration of DINP from polyvinyl Chloride (PVC) Children's Products," U.S. Consumer Product Safety Commission. Bethesda, MD.

Efron, B., and Tibshirani, R., (193), An Introduction to the Bootstrap. Chapman and Hall, New York.

Greene, M.A. (1998), "Statistical Analysis for Prediction of DINP Intake by Young Children." U.S. Consumer Product Safety Commission. Bethesda, MD.

Groot, M. E., Lekkerkerk, M, C., and Steenbekkers, L. P. A., (1998), "Mouthing behavior of young children," Agricultural University Wageningen, Wageningen, The Netherlands.

Hahn, G.J. and Meeker, W.Q. (1991) *Statistical Intervals: A Guide for Practitioners.* John Wiley and Sons, NY.

Koneman, W. H., ed (1998), "Phthalate release from soft PVC Baby Toys: Report from the Dutch Consensus Group," RIVM report 613320 002, Bilthoven, Netherlands.

Snyder, R. G., Schneider, L. W., Owings, C. L., Reynolds, H. M., Golomb, H., and Schork, M. A., (1977), "Anthropometry of Infants, Children and Youths to Age 18 for Product Safety Design," Final Report prepared for the U. S. Consumer Product Safety Commission. University of Michigan, Ann Arbor, MI.

Van Veen, M. P. (1998), "Exposure Assessment (Summary Report)", Annex 4, in W. H. Konemann (ed), "Phthalate release from soft PVC Baby Toys: Report from the Dutch Consensus Group," RIVM report 613320 002, Bilthoven, Netherlands.